

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION  
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété  
Intellectuelle  
Bureau international



(43) Date de la publication internationale  
28 juillet 2005 (28.07.2005)

PCT

(10) Numéro de publication internationale  
**WO 2005/069272 A1**

(51) Classification internationale des brevets<sup>7</sup> : G10H 1/00

(21) Numéro de la demande internationale :  
PCT/FR2003/003730

(22) Date de dépôt international :  
15 décembre 2003 (15.12.2003)

(25) Langue de dépôt : français

(26) Langue de publication : français

(71) Déposant (pour tous les États désignés sauf US) :  
FRANCE TELECOM [FR/FR]; 6, place d'Alleray,  
F-75015 Paris (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : NICOL,  
Rozenn [FR/FR]; 2, Place du Martray, f-22450 La Roche  
Derrien (FR). VIRETTE, David [FR/FR]; 25, Avenue  
Ernest Renan, f-22300 Lannion (FR). EMERIT, Marc  
[FR/FR]; 10, cité d'Aleth, f-35000 Rennes (FR).

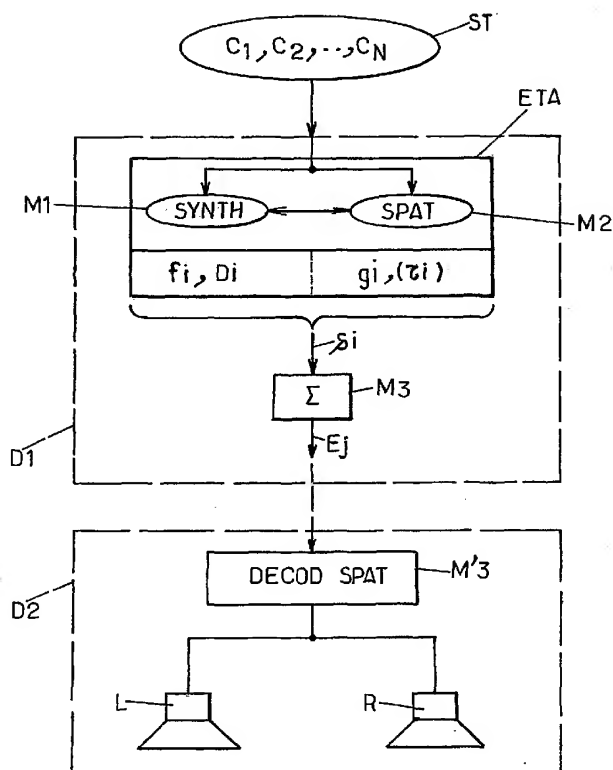
(74) Mandataires : VERDURE, Stéphane, etc.; Cabinet  
Plasseraud, 65/67 rue de la Victoire, F-75440 Paris Cedex  
09 (FR).

(81) États désignés (national) : AE, AG, AL, AM, AT, AU, AZ,  
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ,  
DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM,  
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,  
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,  
MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD,

[Suite sur la page suivante]

(54) Title: METHOD FOR SYNTHESIZING ACOUSTIC SPATIALIZATION

(54) Titre : PROCEDE DE SYNTHÈSE ET DE SPATIALISATION SONORES



(57) Abstract: The invention relates to the synthesis and the joint spatialization of sounds emitted by virtual sources. According to the invention, a step (ETA) is provided that consists of determining parameters including at least one gain ( $g_i$ ) for defining, at the same time, a loudness characterizing the nature of the virtual source and the position of the source relative to a predetermined origin.

(57) Abrégé : L'invention concerne la synthèse et la spatialisation conjointe de sons issus de sources virtuelles. Selon l'invention, on prévoit une même étape (ETA) consistant à déterminer des paramètres incluant au moins un gain ( $g_i$ ), pour définir en même temps: une intensité sonore caractérisant la nature de source virtuelle, et la position de la source par rapport à une origine prédéterminée.

WO 2005/069272 A1



SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG,  
US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **États désignés (régional)** : brevet ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Déclaration en vertu de la règle 4.17 :**

— *relative à la qualité d'inventeur (règle 4.17.iv)) pour US seulement*

**Publiée :**

— *avec rapport de recherche internationale*

*En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.*

Procédé de synthèse et de spatialisation sonores

La présente invention concerne la synthèse de signaux audio, notamment dans des applications d'édition musicale, de jeux vidéo, ou encore de sonneries pour des téléphones mobiles.

Plus particulièrement, l'invention concerne à la fois des techniques de synthèse sonore et des techniques du son tridimensionnel (ou "3D").

Pour offrir des services innovants, basés sur la synthèse sonore (pour créer des sonneries, ou encore dans le cadre de jeux sur téléphones mobiles), on cherche actuellement à enrichir les méthodes de synthèse sonore. Cependant, les terminaux étant limités en termes de mémoire et de puissance de calcul, il est préférable de développer des procédés à la fois efficaces et économes en complexité.

**\* Les techniques de synthèse sonore**

De nombreuses techniques de synthèse sonore ont été développées dans les dernières décennies. On indique qu'il n'existe pas, en réalité, de technique universelle capable de générer n'importe quel son. En effet, les modèles de production qui ont été établis jusqu'à présent ont chacun leurs restrictions. Une taxinomie établie par Julius Smith dans :

*"Viewpoints on the History of Digital Synthesis"*, Smith J.O; Keynote paper, Proc. Int. Comp. Music Conf. 1991, Montreal, est présentée ci-dessous.

Les techniques sont catégorisées en quatre groupes :

- les techniques calculatoires (FM pour "*frequency modulation*", "*waveshaping*" pour le travail de formes d'ondes, etc),
- le "*sampling*" et autres traitements d'enregistrements (par exemple une synthèse par table d'onde, etc),
- les techniques fondées sur des modèles spectraux (tels que la synthèse additive, ou encore le "*source-filtre*", etc),
- les techniques fondées sur des modèles physiques (synthèse modale, par guides d'onde,...).

Certaines techniques, selon leur usage, peuvent tomber dans plusieurs catégories.

Le choix de la technique de synthèse adaptée à un terminal ou à un système de rendu peut se faire en fonction de trois familles de critères, notamment des critères du type de ceux proposés par le laboratoire d'acoustique et de traitement du signal de l'université d'Helsinki dans le cadre d'une évaluation des différentes méthodes de synthèse:

**"*Evaluation of Modern Sound Synthesis Methods*"**, Tolonen T., Välimäki V., Karjalainen M; Report 48, Espoo 1998.

Une première famille de critères concerne l'emploi des paramètres suivants :

- l'intuitivité,
- la perceptibilité,
- le sens physique

- et le comportement.

La qualité et la diversité des sons qui sont produits déterminent la seconde famille de critères, selon les  
5 paramètres suivants :

- robustesse de l'identité du son,
- étendue de la palette sonore, et
- avec une phase préliminaire d'analyse, le cas échéant.

10

Enfin, la troisième famille de critères traite des solutions d'implémentation, avec des paramètres tels que :

- le coût des calculs,
- la mémoire nécessaire,
- 15 - le contrôle, la latence et les traitements multitâches.

Il est apparu récemment que les techniques reposant sur une modélisation spectrale (avec reproduction de l'image  
20 spectrale perçue par un auditeur) ou une modélisation physique (avec simulation de l'origine physique du son) sont les plus satisfaisantes et présentent un large potentiel pour les systèmes futurs.

25 Cependant, actuellement, les méthodes basées sur la synthèse par table d'ondes sont les plus répandues. Le principe de cette technique est le suivant. Tout d'abord, tous les signaux audio naturels peuvent se décomposer en quatre phases : attaque, déclin, soutien et relâchement,  
30 regroupées généralement sous le terme "d'enveloppe ADSR"

(de l'anglais: Attack, Decay, Sustain, Release) qui sera décrite plus loin.

Le principe de la synthèse par table d'ondes consiste à  
5 prendre une ou plusieurs périodes de signal (correspondant à un enregistrement ou à un signal synthétique), puis à lui appliquer des traitements (avec mise en boucle, modification de la fréquence fondamentale; etc) et enfin à lui appliquer l'enveloppe ADSR précitée. Cette méthode de  
10 synthèse très simple permet d'obtenir des résultats satisfaisants. Une technique voisine de la synthèse par table d'ondes est celle dite de "sampling" qui s'en distingue toutefois par le fait qu'elle utilise des enregistrements de signaux naturels à la place de signaux  
15 synthétiques.

Un autre exemple de synthèse simple est la synthèse par modulation de fréquence, plus connue sous le nom de "synthèse FM". Ici, on effectue une modulation de  
20 fréquence pour laquelle la fréquence du modulant et du modulé ( $f_m$  et  $f_c$ ) se trouve dans le domaine audible (20 à 20000 Hz). On indique aussi que les amplitudes respectives des harmoniques par rapport au mode fondamental peuvent être choisies pour définir un timbre du son.

25

Il existe différents formats de transmission de l'information destinée aux synthétiseurs sonores. Tout d'abord, il est possible de transmettre une partition musicale aux formats MIDI ou selon la norme MPEG4-  
30 Structured Audio pour qu'elle soit ensuite synthétisée par la technique de synthèse sonore choisie. Dans certains

systèmes, il est aussi possible de transmettre des informations sur les instruments à utiliser par le synthétiseur, par exemple en utilisant le format DLS qui permet de transmettre les informations nécessaires à la  
5 synthèse de sons par table d'ondes. De même, des langages algorithmiques de type "CSound" ou "MPEG-4 SAOL" permettent de représenter les sons en temps que technique de synthèse sonore.

10 La présente invention concerne la combinaison de la synthèse sonore avec la spatialisation des sons issus de cette synthèse. On rappelle ci-après quelques techniques de spatialisation sonore connues.

15 **\* Les techniques de spatialisation sonore**

Il s'agit de procédés de traitement du signal audio appliqués à la simulation de phénomènes acoustiques et psycho-acoustiques. Ces techniques visent la génération de signaux à diffuser sur haut-parleurs ou sur écouteurs,  
20 afin de donner à l'auditeur l'illusion auditive de sources sonores placées à une position prédéterminée autour de lui. Elles trouvent une application avantageuse dans la création de sources et d'images sonores virtuelles.

25 Parmi les techniques de spatialisation sonore, on distingue principalement deux catégories.

Les procédés basés sur une approche physique consistent généralement à reproduire le champ sonore à l'identique du  
30 champ acoustique original au sein d'une zone de dimensions finies. Ces procédés ne prennent pas en compte a priori

les propriétés perceptives du système auditif, notamment en termes de localisation auditive. Avec de tels systèmes, l'auditeur est ainsi plongé dans un champ en tout point identique à celui qu'il aurait perçu en présence des sources réelles et il est donc capable de localiser les sources sonores comme dans une situation d'écoute réelle.

Les procédés basés sur une approche psycho-acoustique cherchent plutôt à tirer parti des mécanismes de perception sonore 3D afin de simplifier le processus de reproduction sonore. Par exemple, au lieu de reproduire le champ sonore sur toute une zone, on peut se contenter de le reproduire uniquement au niveau des deux oreilles de l'auditeur. De même, on peut imposer une reproduction fidèle du champ sonore sur une fraction du spectre seulement, afin de relâcher la contrainte sur le reste du spectre. L'objectif est de tenir compte des mécanismes de perception du système auditif afin d'identifier la quantité minimale d'information à reproduire pour obtenir un champ psycho-acoustiquement identique au champ original, c'est à dire tel que l'oreille, en raison de la limitation de ses performances, soit incapable de les distinguer l'un de l'autre.

Dans la première catégorie, différentes techniques ont été identifiées :

- l'holophonie, qui est typiquement une technique de reconstruction physique d'un champ sonore, puisqu'elle constitue l'équivalent acoustique de l'holographie. Elle consiste à reproduire un champ sonore à partir d'un enregistrement sur une surface



(sphère creuse, ou autre). De plus amples détails sont donnés dans :

**"Restitution sonore spatialisée sur une zone étendue: Application à la téléprésence"**, R. Nicol; Thèse de  
5 l'Université du Maine, 1999 ;

- la technique ambiophonique (de l'anglais "ambisonic"), qui est un autre exemple de reconstruction physique du champ acoustique, utilisant une décomposition du champ sonore sur une  
10 base de fonctions propres, appelées "harmoniques sphériques".

Dans la seconde catégorie, on identifie par exemple :

- la stéréophonie, qui exploite des différences de  
15 temps ou d'intensité pour positionner les sources sonores entre deux haut-parleurs, sur la base des différences interaurales de temps et d'intensité qui définissent les critères perceptifs de localisation auditive dans un plan horizontal ;
- les techniques binaurales qui visent à reconstruire  
20 le champ sonore uniquement au niveau des oreilles de l'auditeur, de telle sorte que ses tympans perçoivent un champ sonore identique à celui qu'auraient induit les sources réelles.

25

Chaque technique se caractérise par une méthode spécifique d'encodage et de décodage des informations de spatialisation sous un format adéquat de signaux audio.

30 Les différentes techniques de spatialisation sonore se distinguent également par l'étendue de la spatialisation

qu'elles fournissent. Typiquement, une spatialisation 3D telle que l'encodage ambiophonique, l'holophonie, la synthèse binaurale ou transaurale (qui est une transposition de la technique binaurale sur deux haut-parleurs distants) inclut toutes les directions de l'espace. Par ailleurs, une spatialisation bidimensionnelle ("2D"), telle que la stéréophonie, ou une restriction 2D de l'holophonie ou de la technique ambisonique, est limitée au plan horizontal.

10

Enfin les différentes techniques se distinguent par leurs possibles systèmes de diffusion, par exemple :

- une diffusion sur casque pour les techniques binaurales, ou la stéréophonie,
- 15 - une diffusion sur deux haut-parleurs, notamment pour la stéréophonie ou pour un système transaural,
- ou une diffusion sur un réseau à plus de deux haut-parleurs, pour une zone d'écoute étendue (notamment pour des applications multi-auditeurs), en
- 20 holophonie, ou en restitution ambiophonique.

Une large gamme de dispositifs actuels offre des possibilités de synthèse sonore. Ces dispositifs vont de l'instrument de musique (tel qu'un clavier, une boîte à rythmes, ou autres), des terminaux mobiles, par exemple de type PDA (pour "Personal Digital Assistant"), ou encore des ordinateurs sur lesquels sont installés des logiciels d'édition musicale, ou encore des pédaliers d'effets dotés d'une interface MIDI. Les systèmes de restitution sonore (casque, haut-parleurs stéréo ou systèmes à haut-parleurs multiples) et la qualité des systèmes de synthèse de sons

30

sont très variés, notamment selon les capacités de calcul plus ou moins limitées et selon les environnements d'utilisation de tels systèmes.

5 On connaît actuellement des systèmes capables de spatialiser des sons préalablement synthétisés, en particulier par une mise en cascade d'un moteur de synthèse sonore et d'un moteur de spatialisation. La spatialisation est alors appliquée au signal de sortie du  
10 synthétiseur (sur une voie mono ou deux voies stéréo) après mixage des différentes sources. On connaît ainsi des mises en œuvre de cette solution pour alors spatialiser les sons issus d'un synthétiseur.

15 On connaît plus généralement des mises en œuvre consistant en des moteurs de rendu 3D, qui peuvent s'appliquer à tout type de signaux audionumériques, qu'ils soient synthétiques ou non. Par exemple, les différents instruments de musique d'une partition MIDI (format  
20 classique de synthèse sonore) peuvent être alors positionnés dans l'espace sonore. Toutefois, pour obtenir une telle spatialisation, il faut d'abord convertir les signaux midi en signaux audionumériques et ensuite appliquer à ces derniers un traitement de spatialisation.

25

Cette mise en œuvre est particulièrement coûteuse en termes de temps de traitement et de complexité de traitement.

L'un des buts de la présente invention est un procédé de synthèse sonore offrant la possibilité de spatialiser directement les sons synthétiques.

5 Plus particulièrement, un but de la présente invention est d'associer à la synthèse sonore des outils de spatialisation de qualité satisfaisante. Toutefois, cette association cumule la complexité due à la synthèse sonore à celle de la spatialisation, ce qui rend difficile  
10 l'implémentation d'une synthèse sonore spatialisée sur des terminaux à fortes contraintes (c'est-à-dire à puissance de calcul et à taille mémoire relativement limitées).

Un autre but de la présente invention vise une  
15 optimisation de la complexité de la spatialisation des sons synthétiques selon les capacités du terminal.

A cet effet, la présente invention propose tout d'abord un procédé de synthèse et de spatialisation sonores, dans  
20 lequel un son synthétique à générer est caractérisé par la nature d'une source sonore virtuelle et par sa position par rapport à une origine choisie.

Le procédé au sens de l'invention comporte une étape  
25 conjointe consistant à déterminer des paramètres incluant au moins un gain, pour définir en même temps:

- une intensité sonore caractérisant la nature de la source, et
- la position de la source par rapport à une origine  
30 prédéterminée.

On comprendra ainsi que la présente invention permet d'intégrer une technique de spatialisation du son à une technique de synthèse sonore, de façon à obtenir un traitement global utilisant des paramètres communs pour la  
5 mise en œuvre des deux techniques.

Dans une réalisation, la spatialisation de la source virtuelle s'effectue en contexte ambiophonique. Le procédé comporte alors une étape de calcul de gains associés à des  
10 composantes ambiophoniques dans une base d'harmoniques sphériques.

Dans une variante, le son synthétique est destiné à être restitué en contexte holophonique, ou binaural, ou  
15 transaural, sur une pluralité de voies de restitution. On comprendra en particulier que cette "*pluralité de voies de restitution*" peut aussi bien concerner deux voies de restitution, en contexte binaural ou transaural, ou encore plus de deux voies de restitution, par exemple en contexte  
20 holophonique. Pendant ladite étape conjointe, on détermine en outre un retard entre voies de restitution, pour définir en même temps :

- un instant de déclenchement du son caractérisant la nature de la source, et
- 25 - la position de la source par rapport à une origine prédéterminée.

Dans cette réalisation, la nature de la source virtuelle est paramétrée au moins par une variation temporelle d'intensité sonore, sur une durée choisie et incluant un  
30 instant de déclenchement du son. En pratique, cette

variation temporelle peut avantageusement être représentée par une enveloppe ADSR telle que décrite ci-avant.

Préférentiellement, cette variation comporte au moins :

- 5 - une phase d'attaque instrumentale,
- une phase de déclin,
- une phase de soutien, et
- une phase de relâchement.

Bien entendu, des variations plus complexes d'enveloppe  
10 peuvent être envisagées.

La spatialisation de la source virtuelle s'effectue  
préférentiellement par une synthèse binaurale basée sur  
une décomposition linéaire de fonctions de transfert, ces  
15 fonctions de transfert s'exprimant par une combinaison  
linéaire de termes dépendant de la fréquence du son et  
pondérés par des termes dépendant de la direction du son.  
Cette mesure s'avère avantageuse notamment lorsque la  
position de la source virtuelle est susceptible d'évoluer  
20 dans le temps et/ou lorsque plusieurs sources virtuelles  
sont à spatialiser.

Préférentiellement, la direction est définie par au moins  
un angle azimutal (pour une spatialisation dans un simple  
25 plan) et, de préférence, par un angle azimutal et un angle  
d'élévation (pour une spatialisation en trois dimensions).

Dans le contexte d'une synthèse binaurale basée sur une  
décomposition linéaire des fonctions de transfert, la  
30 position de la source virtuelle est avantageusement  
paramétrée au moins par :

- plusieurs filtrages, fonctions de la fréquence sonore,
- plusieurs gains de pondération associés chacun à un filtrage, et
- 5     - un retard par voie "gauche" et "droite".

Préférentiellement, la nature de la source virtuelle est paramétrée au moins par un timbre sonore, en associant des intensités sonores relatives choisies à des harmoniques  
10 d'une fréquence correspondant à une hauteur tonale du son. En pratique, cette modélisation s'effectue avantageusement par une synthèse FM, décrite ci-avant.

Dans une réalisation avantageuse, on prévoit un moteur de  
15 synthèse sonore propre à générer des sons spatialisés, par rapport à une origine prédéterminée.

Préférentiellement, le moteur de synthèse est implémenté en contexte d'édition musicale, et l'on prévoit en outre  
20 une interface homme/machine pour placer la source virtuelle à une position choisie par rapport à l'origine prédéterminée.

Pour synthétiser et spatialiser une pluralité de sources  
25 virtuelles, on assigne chaque source à une position respective, préférentiellement en utilisant une décomposition linéaire des fonctions de transfert en contexte binaural, comme indiqué ci-avant.

30 La présente invention vise aussi un module de génération de sons synthétiques, comprenant notamment un processeur,

et comportant en particulier une mémoire de travail propre à stocker des instructions pour la mise en œuvre du procédé ci-avant, de manière à traiter simultanément une synthèse et une spatialisation du son, selon l'un des  
5 avantages que procure la présente invention.

A ce titre la présente invention vise aussi un produit programme d'ordinateur, stocké dans une mémoire d'une unité centrale ou d'un terminal, notamment mobile, ou sur  
10 un support amovible propre à coopérer avec un lecteur de ladite unité centrale, et comportant des instructions pour la mise en œuvre du procédé ci-avant.

D'autres caractéristiques et avantages de l'invention apparaîtront à l'examen de la description détaillée ci-après, et des dessins annexés sur lesquels :

- la figure 1 illustre schématiquement des positions de sources sonores  $i$  et des positions de microphones  $j$  dans l'espace tridimensionnel,
- 20 - la figure 2 représente schématiquement un traitement simultané de spatialisation et de synthèse sonore, au sens de l'invention,
- la figure 3 représente schématiquement l'application de fonctions de transfert HRTFs à des signaux  $S_i$  pour une spatialisation en synthèse binaurale ou transaurale,
- 25 - la figure 4 représente schématiquement l'application d'une paire de retards (un retard par voie gauche ou droite) et de plusieurs gains (un gain par filtre directionnel) en synthèse binaurale ou transaurale,
- 30 utilisant la décomposition linéaire des HRTFs,



- la figure 5 représente schématiquement l'intégration du traitement de spatialisation, au sein d'une pluralité de générateurs de sons synthétiques, pour une spatialisation et une synthèse sonore en une seule étape,
- 5 - la figure 6 représente un modèle d'enveloppe ADSR en synthèse sonore,
- et la figure 7 représente schématiquement un générateur de son en synthèse FM.

10 On rappelle que la présente invention propose d'intégrer une technique de spatialisation du son à une technique de synthèse sonore de façon à obtenir un traitement global, optimisé, de synthèse sonore spatialisée. Dans le contexte de terminaux à fortes contraintes, la mise en commun de  
15 certaines des opérations de synthèse sonore, d'une part, et de spatialisation sonore, d'autre part, s'avère particulièrement intéressante.

De façon générale, un moteur de synthèse sonore  
20 (typiquement un "*synthétiseur*") a pour rôle de générer un ou plusieurs signaux synthétiques, sur la base d'un modèle de synthèse des sons, modèle qui est piloté à partir d'un ensemble de paramètres, appelés ci-après "*paramètres de synthèse*". Les signaux synthétiques générés par le moteur  
25 de synthèse peuvent correspondre à des sources sonores distinctes (qui sont, par exemple, les différents instruments d'une partition) ou peuvent être associés à une même source, par exemple dans le cas de différentes notes d'un même instrument. Par la suite, les termes  
30 "*générateur de sons*" désignent un module de production d'une note musicale. Ainsi, on comprendra qu'un

synthétiseur est composé d'un ensemble de générateurs de sons.

De façon générale encore, un outil de spatialisation  
5 sonore est un outil qui admet un nombre donné de signaux audio en entrée, ces signaux étant représentatifs de sources sonores et, en principe, exempts de traitement de spatialisation. On indique en fait que, si ces signaux ont déjà subi un traitement spatialisé, ce traitement  
10 préalable n'est pas pris en compte ici. L'outil de spatialisation a pour rôle de traiter les signaux d'entrée, selon un schéma qui est spécifique à la technique de spatialisation choisie, pour générer un nombre donné de signaux de sortie qui définissent les  
15 signaux spatialisés représentatifs de la scène sonore au format de spatialisation choisi. La nature et la complexité du traitement de spatialisation dépendent évidemment de la technique choisie, selon que l'on considère un rendu au format stéréophonique, binaural,  
20 holophonique ou ambiophonique.

Plus particulièrement, pour de nombreuses techniques de spatialisation, il apparaît que le traitement se ramène à une phase d'encodage et une phase de décodage, comme on le  
25 verra plus loin.

L'encodage correspond à la prise de son du champ sonore généré par les différentes sources à un instant donné. Ce système de prise de son "virtuel" peut être plus ou moins  
30 complexe suivant la technique de spatialisation sonore retenue. Ainsi, on simule une prise de son par un nombre

plus ou moins important de microphones avec des positions et des directivités différentes. Dans tous les cas, l'encodage se ramène, pour calculer la contribution d'une source sonore, au moins à l'application de gains et, le plus souvent, de retards (typiquement en holophonie ou en synthèse binaurale ou transaurale) à différentes copies du signal émis par la source. On dispose d'un gain (et, le cas échéant d'un retard) par source pour chaque microphone virtuel. Ce gain (et ce retard) dépendent de la position de la source par rapport au microphone. Si l'on prévoit un système de prise de son virtuel muni de K microphones, on dispose de K signaux en sortie du système d'encodage.

En se référant à la figure 1, le signal  $E_j$  représente la somme des contributions de l'ensemble des sources sonores sur le microphone j. En outre, on appelle :

- $S_i$  le son émis par la source i,
- $E_j$  le signal encodé en sortie du microphone j,
- $G_{ji}$  l'atténuation du son  $S_i$  due à la distance entre la source i et le microphone j, à la directivité de la source, aux obstacles entre la source i et le microphone j, et enfin à la directivité même du microphone j,
- $t_{ji}$  le retard du son  $S_i$  dû à la propagation depuis la source i vers le microphone j, et
- $x, y, z$  les coordonnées cartésiennes de la position de la source, supposée variable dans le temps.

Le signal encodé  $E_j$  est donné par l'expression :

$$E_j(t) = \sum_{i=1}^L \delta(t - t_{ji}(x, y, z)) * G_{ji}(x, y, z) S_i(t)$$

Dans cette expression, on suppose que l'on doit traiter  
L sources ( $i = 1, 2, \dots, L$ ), tandis que le format  
d'encodage prévoit K signaux ( $j = 1, 2, \dots, K$ ). Les gains  
5 et les retards dépendent de la position de la source  $i$  par  
rapport au microphone  $j$  à l'instant  $t$ . L'encodage est donc  
une représentation du champ sonore généré par les sources  
sonores à cet instant  $t$ . On rappelle simplement ici qu'en  
contexte ambiophonique (consistant en une décomposition du  
10 champ dans une base d'harmoniques sphériques), le retard  
n'intervient pas réellement dans le traitement de  
spatialisation.

Dans le cas où les sources sonores sont dans une salle, il  
15 faut ajouter les sources-images. Ce sont les images des  
sources sonores réfléchies par les murs de la salle. Les  
sources-images, en se réfléchissant à leur tour sur les  
murs, génèrent des sources images d'ordres supérieurs.  
Dans l'expression ci-dessus,  $L$  ne représente donc plus le  
20 nombre de sources, mais le nombre de sources auquel on  
additionne le nombre de sources-images. Le nombre de  
sources-images est infini, c'est pourquoi, en pratique, on  
ne garde que les sources images audibles et dont on  
perçoit la direction. Les sources-images qui sont audibles  
25 mais dont on ne perçoit plus la direction sont regroupées  
et leur contribution est synthétisée à l'aide d'un  
réverbérateur artificiel.

L'étape de décodage vise à restituer les signaux  $E_j$   
30 encodés sur un dispositif donné, comportant un nombre  
prédéterminé  $T$  de transducteurs sonores (casque, haut-

parleur). Cette étape consiste à appliquer une matrice  $T \times K$  de filtres aux signaux encodés. Cette matrice dépend uniquement du dispositif de rendu, et non des sources sonores. Suivant la technique d'encodage et de décodage choisie, cette matrice peut être très simple (par exemple l'identité) ou très complexe.

On a représenté schématiquement sur la figure 2 un organigramme reprenant les différentes étapes précitées. Une première étape ST constitue une étape de démarrage au cours de laquelle un utilisateur définit des commandes de sons  $C_1, C_2, \dots, C_N$  à synthétiser et spatialiser (par exemple en prévoyant une interface homme/machine pour définir une note musicale, un instrument pour jouer cette note et une position de cet instrument jouant cette note dans l'espace). En variante, par exemple pour la spatialisation du son auprès d'un terminal mobile, les informations de spatialisation peuvent être transmises dans un flux parallèle au flux audio synthétique, ou encore directement dans le flux audio synthétique.

Ensuite, on indique que l'invention prévoit avantageusement une étape unique ETA de synthèse et spatialisation du son, de façon conjointe. De façon générale, un son peut être défini au moins par :

- la fréquence de son mode fondamental, caractérisant la hauteur tonale,
- sa durée,
- et son intensité.

Ainsi, dans l'exemple d'un synthétiseur à clavier sensitif, si l'utilisateur joue une note *forte*, l'intensité associée à la commande  $C_i$  sera plus grande que l'intensité associée à une note *piano*. Plus  
5 particulièrement, on indique que le paramètre d'intensité peut, de façon générale, prendre en compte le gain de spatialisation  $g_i$  dans un contexte de traitement de spatialisation, comme on le verra ci-après, selon l'un des avantages majeurs que procure la présente invention.

10 En outre, un son est, bien entendu, défini aussi par son instant de déclenchement. Typiquement, si la technique de spatialisation choisie n'est pas un traitement ambiophonique, mais plutôt la synthèse binaurale ou  
15 transaurale, l'holophonie, ou autre, le retard de spatialisation  $\tau_i$  (qui sera décrit en détail plus loin) peut permettre de contrôler en outre l'instant de déclenchement du son.

20 En se référant à nouveau à la figure 2, un dispositif de synthèse et de spatialisation sonore D1 comporte :

- un module de synthèse proprement dit M1, capable de définir, en fonction d'une commande  $C_i$ , au moins la fréquence  $f_i$  et la durée  $D_i$  du son  $i$  associé à cette  
25 commande  $C_i$ , et
- un module de spatialisation M2, capable de définir au moins le gain  $g_i$  (en contexte ambiophonique notamment) et, en outre, le retard de spatialisation  $\tau_i$ , en holophonie ou synthèse binaurale ou  
30 transaurale.

Comme indiqué ci-avant, ces deux derniers paramètres  $g_i$  et  $\tau_i$  peuvent servir conjointement pour la spatialisation, mais aussi pour la synthèse même du son, lorsque l'on définit une intensité sonore (ou un panoramique en stéréophonie) et un instant de déclenchement du son.

Plus généralement, on indique que dans un mode de réalisation préféré, les deux modules M1 et M2 sont regroupés dans un même module permettant de définir en une même étape tous les paramètres du signal  $s_i$  à synthétiser et spatialiser : sa fréquence, sa durée, son gain de spatialisation, son retard de spatialisation, notamment.

Ces paramètres sont ensuite appliqués à un module d'encodage M3 du dispositif de synthèse et de spatialisation sonore D1. Typiquement, par exemple en synthèse binaurale ou transaurale, ce module M3 réalise une combinaison linéaire sur les signaux  $s_i$  qui implique notamment les gains de spatialisation, comme on le verra plus loin. Ce module d'encodage M3 peut en outre appliquer un codage en compression sur les signaux  $s_i$  pour préparer une transmission des données codées vers un dispositif de restitution D2.

On indique toutefois que ce module d'encodage M3 est, dans un mode de réalisation préféré, directement intégré aux modules M1 et M2 ci-avant, de manière à créer directement, au sein d'un module unique D1 qui consisterait simplement en un moteur de synthèse et de spatialisation sonore, les signaux  $E_j$  comme s'ils étaient délivrés par des microphones  $j$ , comme expliqué ci-avant.

Ainsi, le moteur de synthèse et de spatialisation sonore D1 produit, en sortie, K signaux sonores  $E_j$  représentant l'encodage du champ sonore virtuel qu'auraient créé les  
5 différentes sources synthétiques si elles avaient été réelles. A ce stade, on possède une description d'une scène sonore dans un format d'encodage donné.

Bien entendu, il peut être prévu en outre d'ajouter  
10 (ou "de mixer") à cette scène sonore d'autres scènes provenant d'une prise de son réelle ou de la sortie d'autres modules de traitement de sons, à condition qu'elles soient au même format de spatialisation. Le mixage de ces différentes scènes passe ensuite dans un  
15 même et unique système de décodage M'3, prévu en entrée d'un dispositif de restitution D2. Dans l'exemple représenté sur la figure 2, ce dispositif de restitution D2 comprend deux voies, ici pour une restitution binaurale (restitution sur casque stéréophonique) ou transaurale  
20 (restitution sur deux haut-parleurs) sur deux voies L et R.

On décrit ci-après un mode de réalisation préféré de l'invention, ici appliqué à un terminal mobile et dans le  
25 contexte d'une spatialisation sonore par synthèse binaurale.

Sur des terminaux de télécommunication, notamment mobiles, on prévoit naturellement un rendu sonore avec un casque  
30 d'écoute stéréophonique. La technique de positionnement des sources sonores préférée est alors la synthèse



binaurale. Elle consiste, pour chaque source sonore, à filtrer le signal monophonique par des fonctions de transfert acoustiques appelées HRTFs (pour "*Head Related Transfer Functions*" en anglais), qui modélisent les transformations engendrées par le torse, la tête et le pavillon de l'auditeur au signal provenant d'une source sonore. Pour chaque position de l'espace, on peut mesurer une paire de ces fonctions (une fonction pour l'oreille droite, une fonction pour l'oreille gauche). Les HRTFs sont donc fonctions de la position  $[\theta, \phi]$  (où  $\theta$  représente l'azimut et  $\phi$  l'élévation) et de la fréquence sonore  $f$ . On obtient alors, pour un sujet donné, une base de données de  $2M$  fonctions de transfert acoustiques représentant chaque position de l'espace pour chaque oreille ( $M$  étant le nombre de directions mesurées). De manière classique, la mise en œuvre de cette technique se fait sous forme dite "*bicanale*".

Une autre synthèse binaurale, basée sur une décomposition linéaire des HRTFs, correspond à une mise en œuvre qui s'avère plus efficace en particulier lorsque plusieurs sources sonores sont spatialisées, ou dans le cas où les sources sonores changent de position dans le temps. Dans ce cas, on parle de "*synthèse binaurale dynamique*".

25

Ces deux modes de réalisation de la synthèse binaurale sont décrits ci-après.

**\* Synthèse binaurale "bicanale"**

En se référant à la figure 3, la synthèse binaurale bicanale consiste à filtrer le signal de chaque source  $S_i$  ( $i = 1, 2, \dots, N$ ) que l'on souhaite positionner dans l'espace à une position  $[\theta_i, \phi_i]$ , par les fonctions de transfert acoustiques gauche (HRTF\_l) et droite (HRTF\_r) correspondant aux directions appropriées  $[\theta_i, \phi_i]$  (étape 31). On obtient deux signaux qui sont alors additionnés aux signaux gauches et droits résultant de la spatialisation des autres sources (étape 32), pour donner les signaux L et R diffusés aux oreilles gauche et droite du sujet avec un casque d'écoute stéréophonique.

On indique que, dans cette mise en œuvre, les positions des sources sonores ne sont pas sensées évoluer dans le temps. Toutefois, si l'on souhaite faire varier les positions des sources sonores dans l'espace au cours du temps, il est préférable de modifier les filtres utilisés pour modéliser les HRTFs gauche et droite. En revanche, ces filtres se présentant sous la forme de filtres, soit à réponse impulsionnelle finie (FIR), soit à réponse impulsionnelle infinie (IIR), des problèmes de discontinuités des signaux de sortie gauche et droit apparaissent, entraînant des "clicks" audibles. La solution technique employée pour pallier ce problème est de faire tourner deux jeux de filtres binauraux en parallèle. Le premier jeu simule la première position  $[\theta_1, \phi_1]$  à un instant  $t_1$ , le second la seconde position  $[\theta_2, \phi_2]$  à un instant  $t_2$ . Le signal donnant l'illusion d'un déplacement entre les première et seconde positions est alors obtenu par un fondu enchaîné des signaux gauche et

droit résultant des premier et second processus de filtrage. Ainsi, la complexité du système de positionnement des sources sonores est alors multipliée par deux par rapport au cas statique. De plus, le nombre  
5 de filtres à mettre en oeuvre est proportionnel aux nombres de sources à spatialiser.

Si  $N$  sources sonores sont considérées, le nombre de filtres nécessaires est alors de  $2.N$  pour une synthèse  
10 binaurale statique et de  $4.N$  pour une synthèse binaurale dynamique.

On décrit ci-après une variante avantageuse.

15 **\* Synthèse binaurale basée sur une décomposition linéaire des HRTFs**

On indique d'abord qu'une telle mise en œuvre présente une complexité qui ne dépend plus du nombre total de sources à positionner dans l'espace. En effet, ces techniques  
20 permettent de décomposer les HRTFs sur une base de fonctions orthogonales, communes à toutes les positions de l'espace, et ne dépendent donc plus que de la fréquence  $f$ . On réduit ainsi le nombre de filtres nécessaires. Plus particulièrement, le nombre de filtres est fixe et ne  
25 dépend plus du nombre de sources à positionner, de sorte que l'ajout d'une source sonore supplémentaire ne nécessite que l'application d'un retard, avec ensuite une opération de multiplication par plusieurs gains dépendant uniquement de la position  $[\theta, \phi]$  et une opération  
30 d'addition, comme on le verra en référence à la figure 4. Ces techniques de décomposition linéaire ont aussi un

intérêt dans le cas de la synthèse binaurale dynamique (position de sources sonores variable dans le temps). En effet, dans ce cas, on ne fait plus varier les coefficients des filtres, mais seulement les valeurs des gains qui sont fonction de la position.

La décomposition linéaire des HRTFs vise à séparer les dépendances spatiales et fréquentielles des fonctions de transfert. Au préalable, l'excès de phase des HRTFs est extrait, puis modélisé sous la forme d'un retard pur  $\tau$ . La décomposition linéaire s'applique alors à la composante à phase minimale des HRTFs. Chaque HRTF s'écrit comme une somme de  $P$  fonctions spatiales  $C_j(\theta, \varphi)$  et de filtres de reconstruction  $L_j(f)$  :

$$HRTF(\theta, \varphi, f) = \exp(j2\pi f\tau(\theta, \varphi)) \cdot \sum_{j=1}^P C_j(\theta, \varphi) L_j(f) \quad (1)$$

Le schéma de mise en œuvre de la synthèse binaurale basée sur une décomposition linéaire des HRTFs est illustré sur la figure 4. Les retards interauraux  $\tau_i$  (étape 41) associés aux différentes sources sont d'abord appliqués au signal de chaque source à spatialiser  $S_i$  (avec  $i=1, \dots, N$ ). Le signal de chaque source est ensuite décomposé en  $P$  canaux correspondant aux  $P$  vecteurs de base de la décomposition linéaire. A chacun de ces canaux sont alors appliqués les coefficients directionnels  $C_j(\theta_i, \varphi_i)$  (notés  $C_{i,j}$ ) issus de la décomposition linéaire des HRTFs (étape 42). Ces paramètres de spatialisation  $\tau_i$  et  $C_{i,j}$  ont pour particularité de ne dépendre que de la position  $[\theta_i, \varphi_i]$  où l'on souhaite placer la source. Ils ne dépendent pas de la fréquence sonore. Pour chaque source,

le nombre de ces coefficients correspond au nombre  $P$  des vecteurs de base que l'on a utilisés pour la décomposition linéaire des HRTFs.

- 5 Pour chaque canal, les signaux des  $N$  sources sont ensuite additionnés (étape 43) puis filtrés (étape 44) par le filtre  $L_j(f)$  correspondant au  $j^{\text{ième}}$  vecteur de base.

10 Le même schéma est appliqué séparément pour les canaux droit et gauche. La figure 4 distingue les retards appliqués sur la voie gauche ( $\tau_{Li}$ ) et droite ( $\tau_{Ri}$ ), ainsi que les coefficients directionnels appliqués sur la voie gauche ( $C_{i,j}$ ) et droite ( $D_{i,j}$ ). Finalement, les signaux sommés et filtrés aux étapes 44 et 45 sont sommés à  
15 nouveaux (étape 45 de la figure 4), comme dans l'étape 32 de la figure 3, pour une restitution sur casque stéréophonique. On indique que les étapes 41, 42 et 43 peuvent correspondre à l'encodage spatial proprement dit, pour la synthèse binaurale, tandis que les étapes 44 et 45  
20 peuvent correspondre à un décodage spatial avant restitution, qu'effectuerait le module M'3 de la figure 2, comme décrit ci-avant. En particulier, les signaux issus des sommateurs après l'étape 43 de la figure 4 peuvent être véhiculés via un réseau de communication, pour un  
25 décodage spatial et une restitution auprès d'un terminal mobile, aux étapes 44 et 45 décrites ci-avant.

L'intérêt de cette mise en œuvre est que, contrairement à la synthèse binaurale "*bicanale*", l'ajout d'une source  
30 supplémentaire ne nécessite pas l'ajout de deux filtres additionnels (de type FIR ou IIR). En d'autres termes, les

$P$  filtres de base sont partagés par toutes les sources présentes. De plus, dans le cas de la synthèse binaurale dynamique, il est possible de faire varier les coefficients  $C_j(\theta_i, \phi_i)$  sans provoquer de clicks audibles en sortie du dispositif. Dans ce cas, seulement 2. $P$  filtres sont nécessaires, alors que 4. $N$  filtres étaient nécessaires pour la mise en œuvre bicanale dynamique décrite ci-avant.

En d'autres termes, les retards  $\tau$  et les gains  $C$  et  $D$ , qui constituent les paramètres de spatialisation et sont spécifiques à chaque source sonore en fonction de sa position, peuvent donc être dissociés des filtres directionnels  $L(f)$  dans la mise en œuvre de la synthèse binaurale basée sur une décomposition linéaire des HRTFs. En conséquence, les filtres directionnels sont communs aux  $N$  sources, indépendamment de leur position, leur nombre ou leur éventuel déplacement. L'application des paramètres de spatialisation représente alors l'encodage spatial, proprement dit, des signaux relatifs aux sources mêmes, tandis que les filtres directionnels réalisent le traitement effectif de décodage spatial, en vue d'une restitution, qui ne dépend plus de la position des sources, mais de la fréquence sonore.

25

En se référant à la figure 5, cette dissociation entre les paramètres de spatialisation et les filtres directionnels est avantageusement exploitée en intégrant l'application du retard et du gain de spatialisation dans le synthétiseur sonore. La synthèse sonore et l'encodage spatial (retards et gains) piloté par l'azimut et

30

l'élévation sont ainsi réalisés simultanément au sein d'un même module tel qu'un générateur de sons, pour chaque signal sonore (ou note, en édition musicale) à générer (étape 51). Le décodage spatial est ensuite pris en charge  
5 par les filtres directionnels  $L_i(f)$ , comme indiqué ci-avant (étape 52).

On décrit maintenant, en référence aux figures 6 et 7, des étapes de la génération de signaux en synthèse sonore. En  
10 particulier, la figure 6 représente les principaux paramètres d'une enveloppe ADSR du type précité, utilisée couramment dans différentes techniques de synthèse sonore. En particulier, la figure 6 représente la variation temporelle de l'enveloppe d'un signal sonore synthétisé,  
15 par exemple une note jouée sur un piano, avec :

- un paramètre d'attaque, modélisé par une rampe montante 61, correspondant par exemple à la durée d'un enfoncement d'un marteau contre une corde de piano,
- 20 - un paramètre de déclin, modélisé par une rampe descendante 62, à forte décroissance, correspondant par exemple à la durée d'un dégagement d'un marteau d'une corde de piano,
- un paramètre de soutien (vibration libre), modélisé  
25 par une rampe légèrement descendante 63, due à l'amortissement acoustique naturel, correspondant par exemple à la durée d'un son d'une touche de piano enfoncée,
- et un paramètre de relâchement, modélisé par une  
30 rampe descendante 64, correspondant par exemple à

l'amortissement acoustique rapide que produit un feutre sur une corde de piano.

Bien entendu, des variations plus complexes d'enveloppe peuvent être envisagées, comportant par exemple plus de  
5 quatre phases.

On indique toutefois que la plupart des sons synthétisés peuvent être modélisés par une variation d'enveloppe comme décrite ci-avant. Préférentiellement, on définit les  
10 paramètres de l'enveloppe ADSR avant d'effectuer les filtrages prévus pour le traitement de spatialisation, en raison des variables de temps mises en jeu.

On comprendra ainsi que le maximum de l'amplitude sonore  
15 (en unité arbitraire sur la figure 6) peut être défini par le traitement de spatialisation, en correspondance alors aux gains  $C_{ij}$  et  $D_{ij}$  précités, pour chaque voie gauche et droite. De même, l'instant de déclenchement du son (début de la rampe 61) peut être défini à travers les retards  $\tau_{Li}$   
20 et  $\tau_{Ri}$ .

On se réfère maintenant à la figure 7 sur laquelle on a représenté un opérateur simple de synthèse sonore par modulation de fréquence ("synthèse FM"). On définit  
25 initialement une fréquence porteuse  $f_c$  (typiquement la fréquence du mode fondamental), qui définit par exemple la tonalité d'une note de musique. On a recours ensuite à un ou plusieurs oscillateurs OSC1 pour définir une ou plusieurs harmoniques  $f_m$  (correspondant en principe à des  
30 fréquences multiples de la fréquence porteuse  $f_c$ ), auxquelles sont associées des intensités relatives  $I_m$ . Par



exemple, les intensités  $I_m$ , par rapport à l'intensité du mode fondamental, sont plus élevées pour un son métallique (tel que celui d'une corde de guitare neuve). De façon générale, la synthèse FM permet de définir le timbre d'un son synthétisé. Les signaux (sinusoïdes) issus du ou des oscillateurs OSC1 sont additionnés au signal tiré de la fréquence porteuse  $f_c$  par le module AD, lequel délivre un signal à un oscillateur de sortie OSC2 qui reçoit en consigne l'amplitude  $A_c$  du son par rapport à la fréquence porteuse  $f_c$ . Là encore, on indique que cette consigne  $A_c$  peut être directement définie par le traitement de spatialisation, à travers les gains C et D (en synthèse binaurale), comme on l'a vu ci-avant. Finalement, l'oscillateur OSC2 délivre un signal  $S'_i$ , auquel on applique ensuite une enveloppe ADSR du type représenté sur la figure 6, ainsi qu'une paire de retards  $\tau_{Li}$  et  $\tau_{Ri}$  et plusieurs gains  $C_{ij}$  et  $D_{ij}$ , respectivement pour chaque voie gauche et droite, comme représenté sur la figure 4, et pour obtenir finalement un signal tel que l'un des signaux que délivrent les générateurs de sons de la figure 5.

On comprendra ainsi qu'une telle mesure permet d'éviter, de façon particulièrement avantageuse, de générer, à partir d'une partition au format MIDI, les sons dans un format standard de restitution audio (par exemple en format "wave") et de les encoder à nouveau pour une spatialisation du son, comme dans les mises en œuvres connues.

La présente invention permet de mettre en œuvre directement et à la fois les étapes de spatialisation et

de synthèse du son. On comprendra en particulier que tout traitement de synthèse sonore, nécessitant la définition d'une intensité (et, le cas échéant, un instant de déclenchement du son), peut être effectué conjointement  
5 avec un traitement de spatialisation, proposant un gain (et, le cas échéant, un retard) par voie de restitution.

De façon générale, un synthétiseur de son fonctionne à partir de la lecture d'une partition qui regroupe les  
10 informations sur les instruments à synthétiser, les instants où l'on doit jouer les sons, la hauteur de ces sons, leur force, etc. Lors de la lecture de cette partition, on associe à chaque son un générateur de sons, comme indiqué ci-avant en référence à la figure 5.

15 On considère d'abord le cas où une même source joue plusieurs notes simultanément. Ces notes, qui proviennent de la même source, sont spatialisées à la même position et donc avec les mêmes paramètres. Il est alors préféré de  
20 regrouper le traitement de spatialisation pour les générateurs de sons associés à la même source. Dans ces conditions, les signaux associés aux notes issues d'une même source sont préférentiellement sommés au préalable de façon à appliquer le traitement de spatialisation  
25 globalement sur le signal résultant, ce qui, d'une part, réduit avantageusement le coût d'implémentation et, d'autre part, garantit avantageusement la cohérence de la scène sonore.

30 En outre, les gains et les retards peuvent être appliqués en tirant partie de la structure du synthétiseur. D'une

part, les retards (voie gauche et voie droite) de spatialisation sont implémentés sous la forme de lignes à retard. D'autre part, dans le cadre du synthétiseur, les retards sont gérés par les instants de déclenchement des générateurs de sons en accord avec la partition. Dans le cadre d'une synthèse sonore spatialisée, les deux approches précédentes (ligne à retard et contrôle de l'instant de déclenchement) sont combinées de façon à optimiser le traitement. On économise donc une ligne à retard par source, en jouant sur les instants de déclenchement des générateurs de sons. A cet effet, on extrait la différence entre les retards de la voie gauche et de la voie de droite pour la spatialisation. On prévoit ensuite d'ajouter le plus faible des deux retards à l'instant de déclenchement du générateur. Il reste ensuite à appliquer la différence de temps entre les voies gauche et droite à une seule des deux voies, sous la forme d'une ligne à retard, étant bien entendu que cette différence de retard peut prendre aussi bien des valeurs positives que négatives.

En ce qui concerne les gains, le paramètre de balance (ou de "panoramique"), qui est typiquement associé au système stéréophonique, n'a plus lieu d'être. Il est donc possible de supprimer les gains associés à la balance. De plus, le paramètre de volume du générateur de son peut être appliqué au niveau des différents gains correspondant à l'encodage spatial, comme décrit ci-avant.

On indique en outre que la présente invention permet d'appliquer la spatialisation sonore, source par source,

du fait que l'outil de spatialisation est intégré au cœur du moteur de synthèse sonore. Tel n'est pas le cas si l'on procède au contraire par simple mise en cascade du moteur de synthèse et de l'outil de spatialisation. Dans ce cas  
5 en effet, on rappelle que la spatialisation ne peut être appliquée que globalement à l'ensemble de la scène sonore.

Selon un autre avantage que procure la présente invention, on peut associer de façon judicieuse les outils de  
10 synthèse et de spatialisation sonores, afin de réaliser une mise en œuvre optimisée d'un moteur de synthèse sonore spatialisée, avec, en particulier, une optimisation de la combinaison des opérations de synthèse et de spatialisation, en tenant compte notamment d'au moins un  
15 gain et/ou d'un retard de spatialisation, ou encore d'un filtre de spatialisation.

Dans le cas où les paramètres de synthèse appliquent déjà l'un ou plusieurs de ces paramètres (gain, retard,  
20 filtre), les paramètres de spatialisation sont avantageusement pris en compte par simple modification des paramètres de synthèse, sans modification du modèle de synthèse lui-même.

25 Par ailleurs, par la simple addition au moteur de synthèse sonore d'un gain et d'un retard, éventuellement complétés par un filtre, une synthèse sonore spatialisée, basée sur différentes techniques possibles de spatialisation, peut être obtenue. Ces techniques de spatialisation (synthèse  
30 binaurale/transaurale, holophonie, ambiophonie, etc) peuvent être de complexité et de performance variables

mais offrent dans l'ensemble une spatialisation beaucoup plus riche et plus complète que la stéréophonie, avec notamment un rendu naturel et particulièrement immersif de la scène sonore. En effet, la spatialisation sonore au  
5 sens de l'invention conserve tout le potentiel d'un rendu sonore tridimensionnel, notamment en terme d'immersion, avec une véritable spatialisation en 3D.

Bien entendu, on peut prévoir en outre une intégration du  
10 traitement de spatialisation et d'effet de salle, sous la forme simplifiée d'au moins un gain et/ou d'un retard (éventuellement complétés de filtres), et d'un réverbérateur artificiel pour la réverbération tardive.

**Revendications**

1. Procédé de synthèse et de spatialisation sonores, dans lequel un son synthétique à générer est caractérisé par la nature d'une source sonore virtuelle et par sa position par rapport à une origine choisie, caractérisé en ce qu'il comporte une étape conjointe consistant à déterminer des paramètres incluant au moins un gain, pour définir en même temps:
- une intensité sonore caractérisant la nature de la source, et
  - la position de la source par rapport à une origine prédéterminée.
2. Procédé selon la revendication 1, dans lequel la spatialisation de la source virtuelle s'effectue en contexte ambiophonique, caractérisé en ce qu'il comporte une étape de calcul de gains associés à des composantes ambiophoniques dans une base d'harmoniques sphériques.
3. Procédé selon la revendication 1, dans lequel le son synthétique est destiné à être restitué en contexte holophonique, ou binaural, ou transaural, sur une pluralité de voies de restitution, caractérisé en ce que, pendant ladite étape conjointe, on détermine en outre un retard entre voies de restitution, pour définir en même temps:
- un instant de déclenchement du son caractérisant la nature de la source, et
  - la position de la source par rapport à une origine prédéterminée.

4. Procédé selon la revendication 3, caractérisé en ce que la nature de la source virtuelle est paramétrée au moins par une variation temporelle d'intensité sonore, sur une  
5 durée choisie et incluant un instant de déclenchement du son.

5. Procédé selon la revendication 4, caractérisé en ce que ladite variation comporte au moins :  
10 - une phase d'attaque instrumentale,  
- une phase de déclin,  
- une phase de soutien, et  
- une phase de relâchement.

6. Procédé selon l'une des revendications 3 à 5, caractérisé en ce que la spatialisation de la source virtuelle s'effectue par une synthèse binaurale basée sur une décomposition linéaire de fonctions de transfert, ces  
15 fonctions de transfert s'exprimant par une combinaison linéaire de termes dépendant de la fréquence du son ( $L(f)$ ) et pondérés par des termes dépendant de la direction du son ( $\tau_R, \tau_L, C, D$ ).  
20

7. Procédé selon la revendication 6, caractérisé en ce que la direction est définie par au moins un angle azimutal  
25 ( $\theta$ ) et, de préférence, par un angle azimutal ( $\theta$ ) et un angle d'élévation ( $\phi$ ).

8. Procédé selon l'une des revendications 6 et 7, caractérisé en ce que la position de la source virtuelle  
30 est paramétrée au moins par :

- plusieurs filtrages, fonctions de la fréquence sonore ( $L_i(f)$ ),
- plusieurs gains de pondération associés chacun à un filtrage, et
- 5       - un retard par voie "gauche" et "droite".

9. Procédé selon l'une des revendications précédentes, caractérisé en ce que la nature de la source virtuelle est paramétrée au moins par un timbre sonore, en associant des  
10       intensités sonores relatives choisies à des harmoniques d'une fréquence correspondant à une hauteur tonale du son.

10. Procédé selon l'une des revendications précédentes, caractérisé en ce qu'il prévoit un moteur de synthèse  
15       sonore propre à générer des sons spatialisés, par rapport à ladite origine prédéterminée.

11. Procédé selon la revendication 10, dans lequel le moteur de synthèse est implémenté en contexte d'édition  
20       musicale, caractérisé en ce que le procédé prévoit en outre une interface homme/machine pour placer la source virtuelle à une position choisie par rapport à l'origine prédéterminée.

25       12. Procédé selon la revendication 11, prise en combinaison avec la revendication 6, dans lequel on prévoit une pluralité de sources virtuelles à synthétiser et spatialiser, caractérisé en ce que l'on assigne chaque source à une position respective.



13. Module de génération de sons synthétiques, comprenant notamment un processeur, caractérisé en ce qu'il comporte en outre une mémoire de travail propre à stocker des instructions pour la mise en œuvre du procédé selon l'une  
5 des revendications précédentes.

14. Produit programme d'ordinateur, stocké dans une mémoire d'une unité centrale ou d'un terminal, notamment mobile, ou sur un support amovible propre à coopérer avec  
10 un lecteur de ladite unité centrale, caractérisé en ce qu'il comporte des instructions pour la mise en œuvre du procédé selon l'une des revendications 1 à 12.

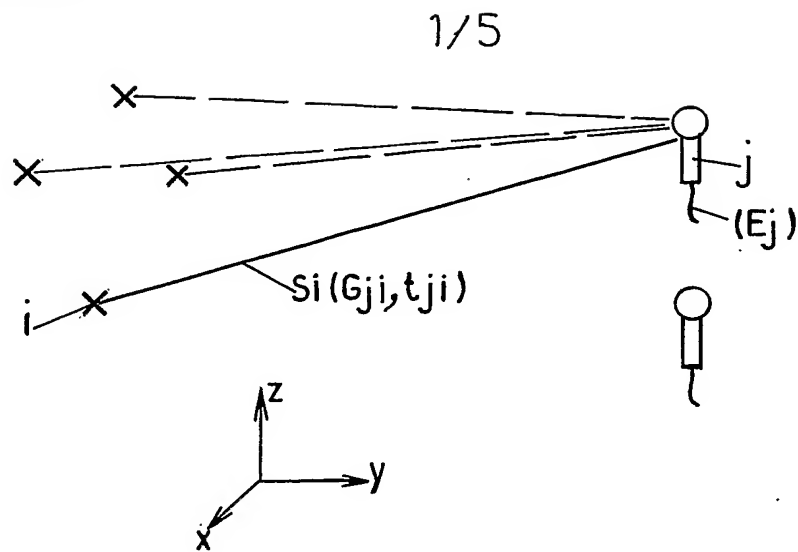
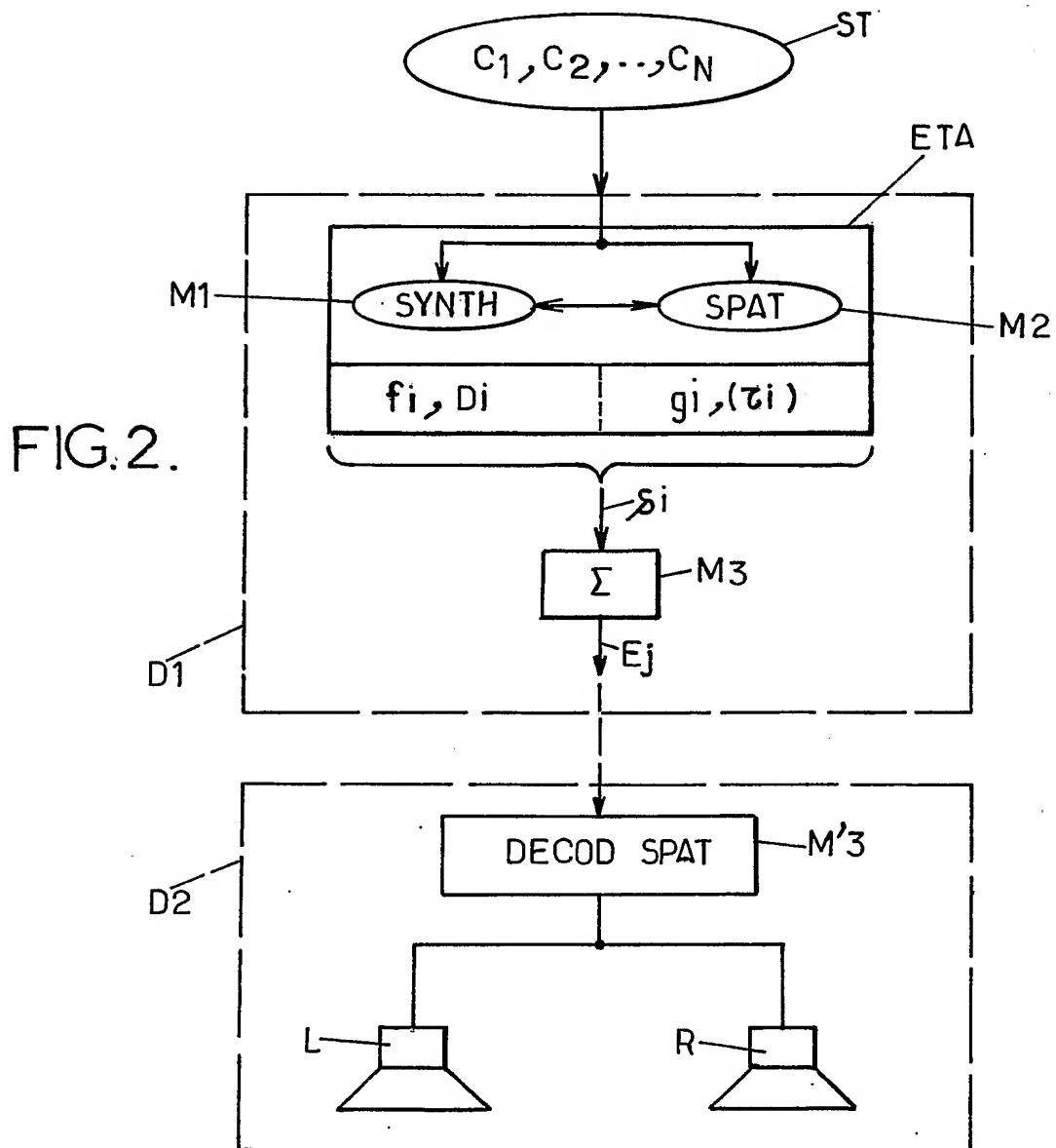


FIG.1.



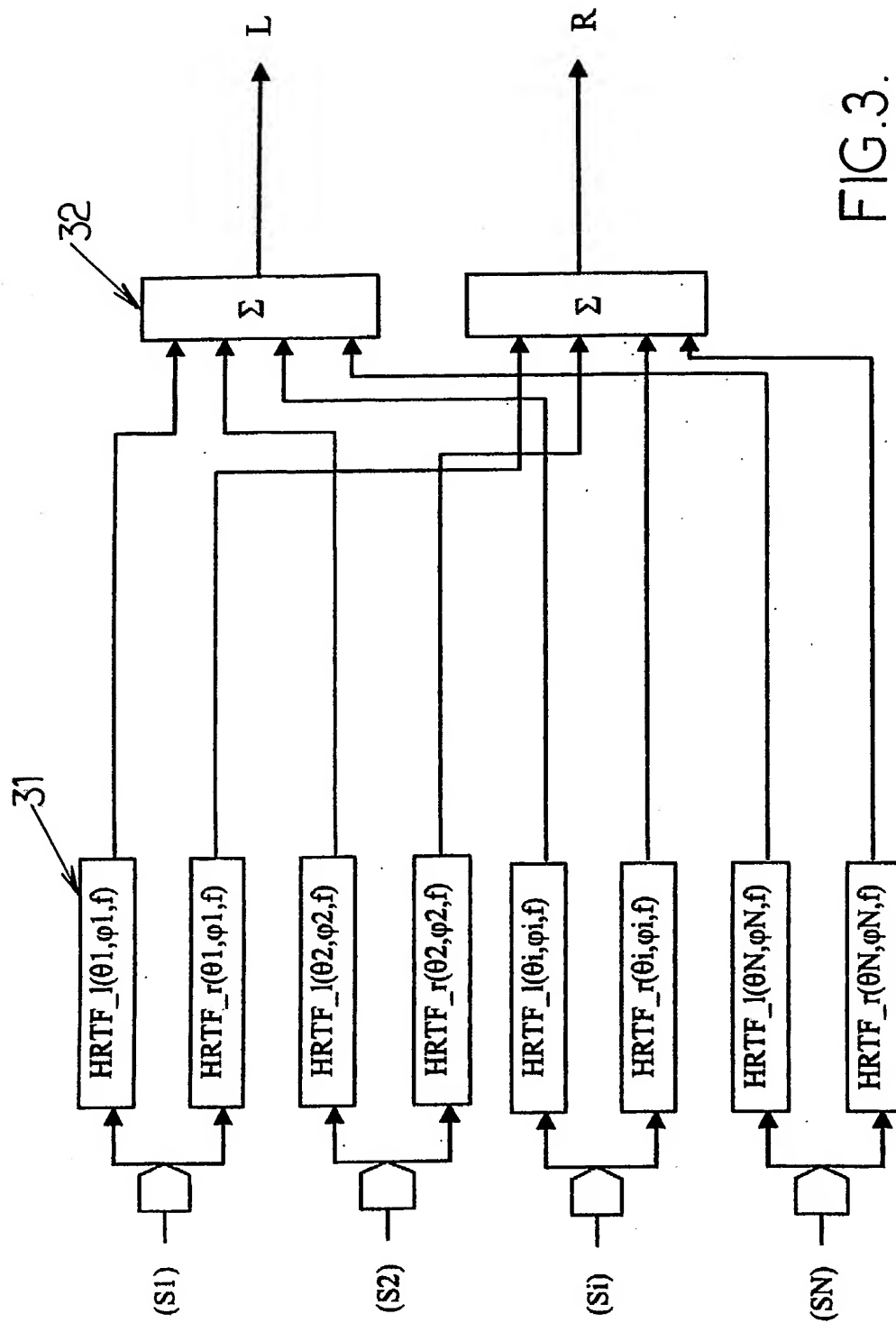


FIG. 3.

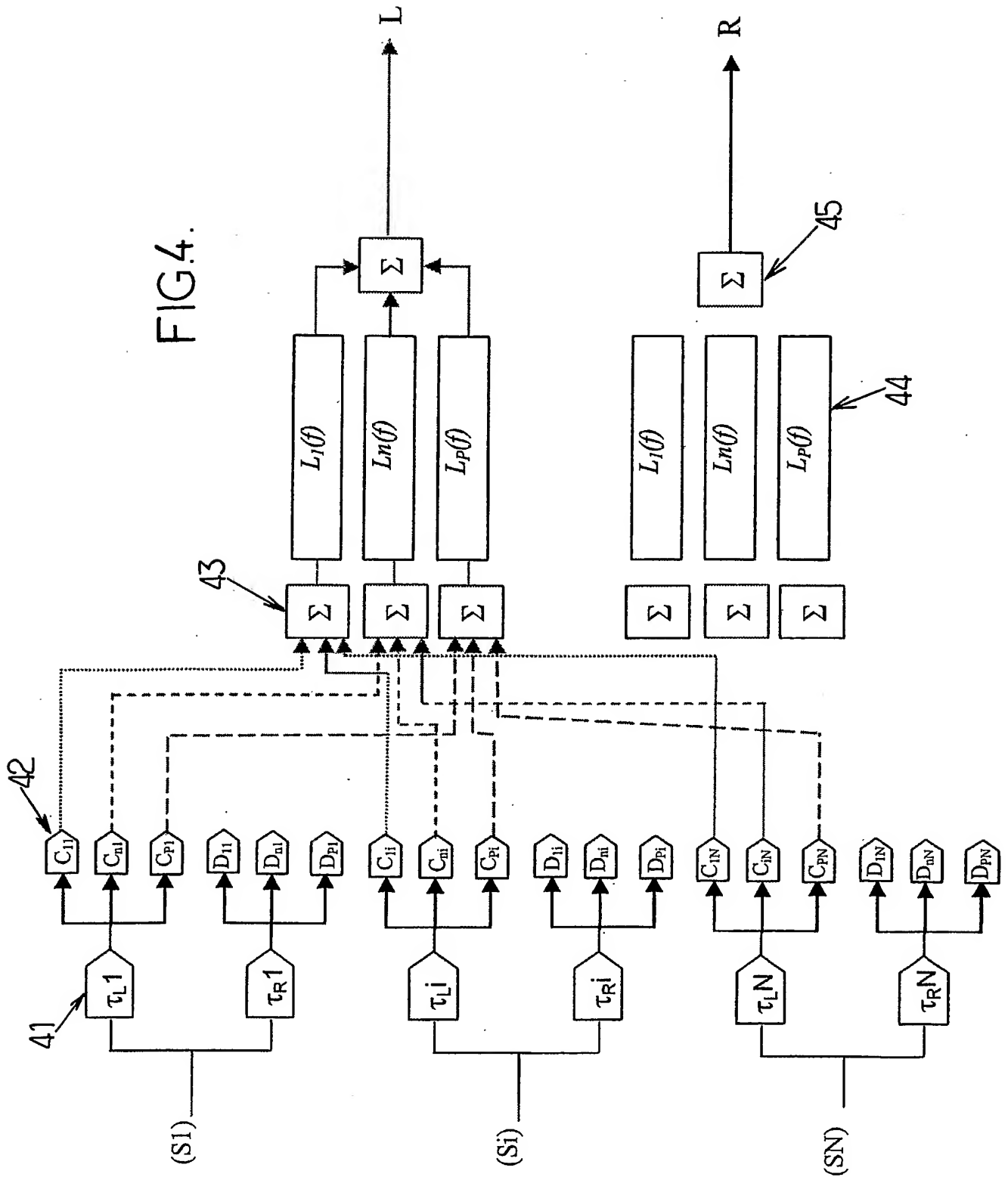
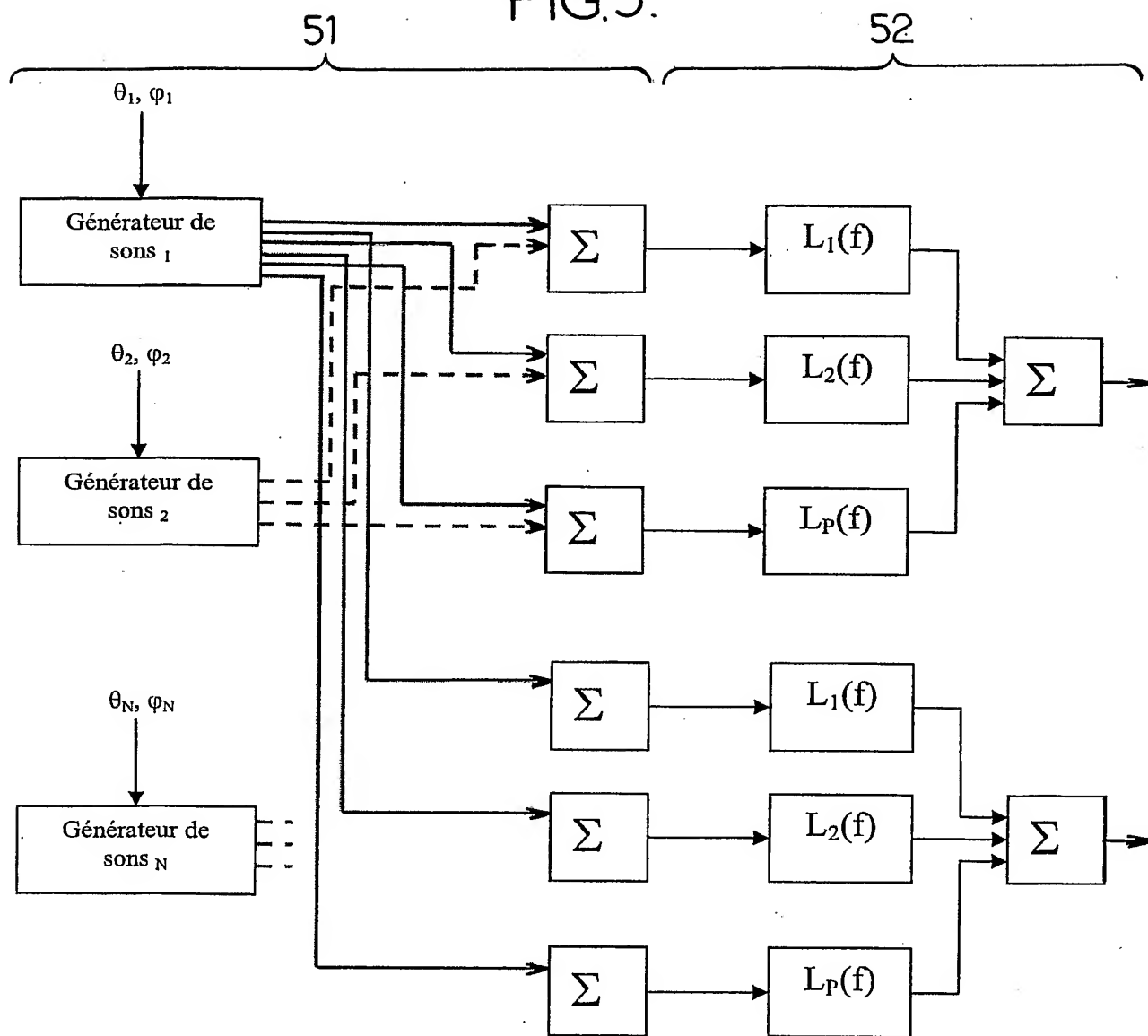


FIG.5.



5/5

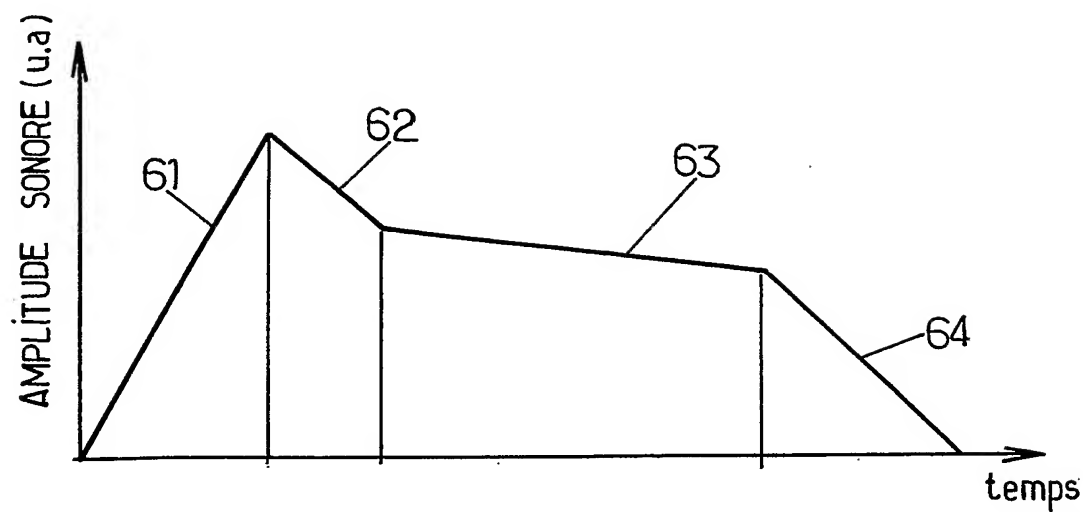
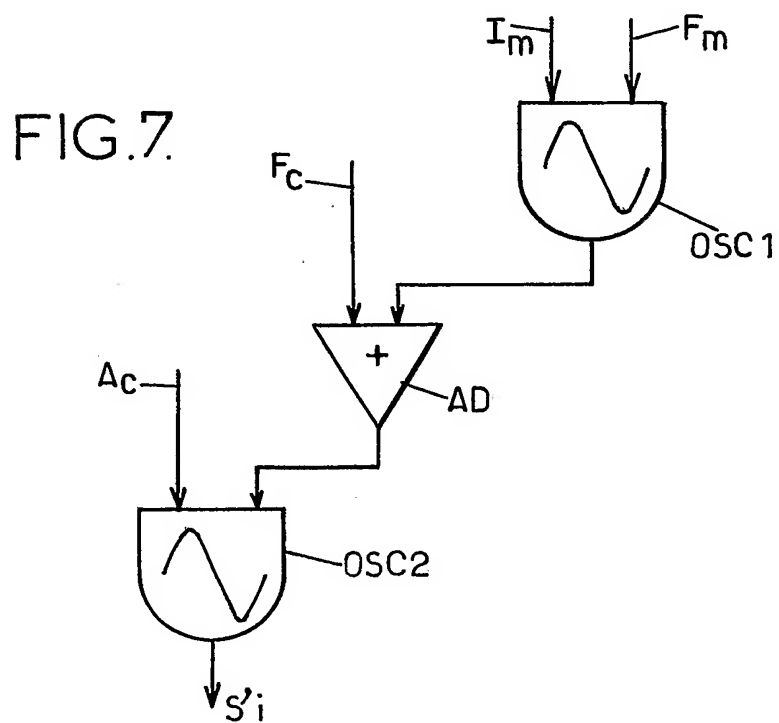


FIG. 6.



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 03/03730

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 G10H1/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 G10H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X A	US 5 977 471 A (ROSENZWEIG MICHAEL D) 2 November 1999 (1999-11-02) column 1, line 65 -column 7, line 32; claims 1,2; figures 1,3 ---	1,3,10, 13,14 2,4-9, 11,12
X A	US 5 822 438 A (SEKINE SATOSHI ET AL) 13 October 1998 (1998-10-13) column 6, line 57 -column 8, line 9; figures 1A,6 ----- -/--	1,3,10, 13,14 2,4-9, 11,12

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

22 April 2004

Date of mailing of the international search report

03/05/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Feron, M

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR 03/03730

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>CHEN J ET AL: "Synthesis of 3D virtual auditory space via a spatial feature extraction and regularization model" VIRTUAL REALITY ANNUAL INTERNATIONAL SYMPOSIUM, 1993., 1993 IEEE SEATTLE, WA, USA 18-22 SEPT. 1993, NEW YORK, NY, USA, IEEE, 18 September 1993 (1993-09-18), pages 188-193, XP010130492 ISBN: 0-7803-1363-1 paragraph '0001!</p> <p>-----</p>	1-14



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 03/03730

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5977471	A	02-11-1999	NONE	
US 5822438	A	13-10-1998	JP 3439485 B2	25-08-2003
			JP 5300597 A	12-11-1993
			JP 2973764 B2	08-11-1999
			JP 6051759 A	25-02-1994
			DE 69322805 D1	11-02-1999
			DE 69322805 T2	26-08-1999
			EP 0563929 A2	06-10-1993
			SG 52771 A1	28-09-1998
			US 5581618 A	03-12-1996

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR 03/03730

**A. CLASSEMENT DE L'OBJET DE LA DEMANDE**  
CIB 7 G10H1/00

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

**B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE**

Documentation minimale consultée (système de classification suivi des symboles de classement)

CIB 7 G10H

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)

EPO-Internal

**C. DOCUMENTS CONSIDERES COMME PERTINENTS**

Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X A	US 5 977 471 A (ROSENZWEIG MICHAEL D) 2 novembre 1999 (1999-11-02) colonne 1, ligne 65 -colonne 7, ligne 32; revendications 1,2; figures 1,3 ---	1,3,10, 13,14 2,4-9, 11,12
X A	US 5 822 438 A (SEKINE SATOSHI ET AL) 13 octobre 1998 (1998-10-13) colonne 6, ligne 57 -colonne 8, ligne 9; figures 1A,6 --- -/--	1,3,10, 13,14 2,4-9, 11,12



Voir la suite du cadre C pour la fin de la liste des documents



Les documents de familles de brevets sont indiqués en annexe

° Catégories spéciales de documents cités:

- \*A\* document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- \*E\* document antérieur, mais publié à la date de dépôt international ou après cette date
- \*L\* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)
- \*O\* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- \*P\* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

\*T\* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

\*X\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

\*Y\* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

\*Z\* document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

22 avril 2004

Date d'expédition du présent rapport de recherche internationale

03/05/2004

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Feron, M

# RAPPORT DE RECHERCHE INTERNATIONALE

Demande Internationale No

PCT/FR 03/03730

## C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	<p>CHEN J ET AL: "Synthesis of 3D virtual auditory space via a spatial feature extraction and regularization model"</p> <p>VIRTUAL REALITY ANNUAL INTERNATIONAL SYMPOSIUM, 1993., 1993 IEEE SEATTLE, WA, USA 18-22 SEPT. 1993, NEW YORK, NY, USA,IEEE,</p> <p>18 septembre 1993 (1993-09-18), pages 188-193, XP010130492</p> <p>ISBN: 0-7803-1363-1</p> <p>alinéa '0001!</p> <p>-----</p>	1-14

# RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande Internationale No

PCT/FR 03/03730

Document brevet cité au rapport de recherche		Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
US 5977471	A	02-11-1999	AUCUN	
US 5822438	A	13-10-1998	JP 3439485 B2	25-08-2003
			JP 5300597 A	12-11-1993
			JP 2973764 B2	08-11-1999
			JP 6051759 A	25-02-1994
			DE 69322805 D1	11-02-1999
			DE 69322805 T2	26-08-1999
			EP 0563929 A2	06-10-1993
			SG 52771 A1	28-09-1998
			US 5581618 A	03-12-1996